Linear Classification and Computation Graphs
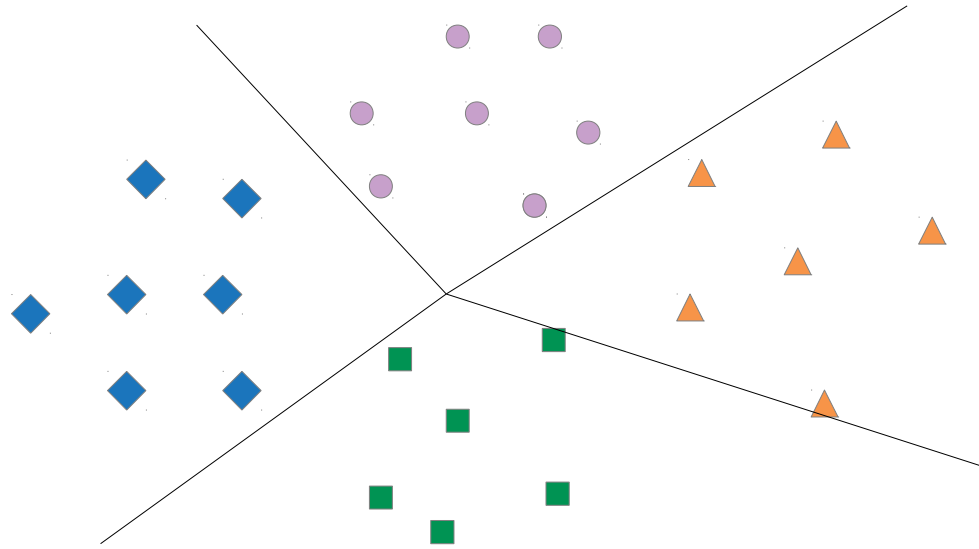
Linear Classification

Linear Classification

- Input: $x \in \mathbb{R}^d$

- Output: $y \in \{1, 2, \ldots, k\}$

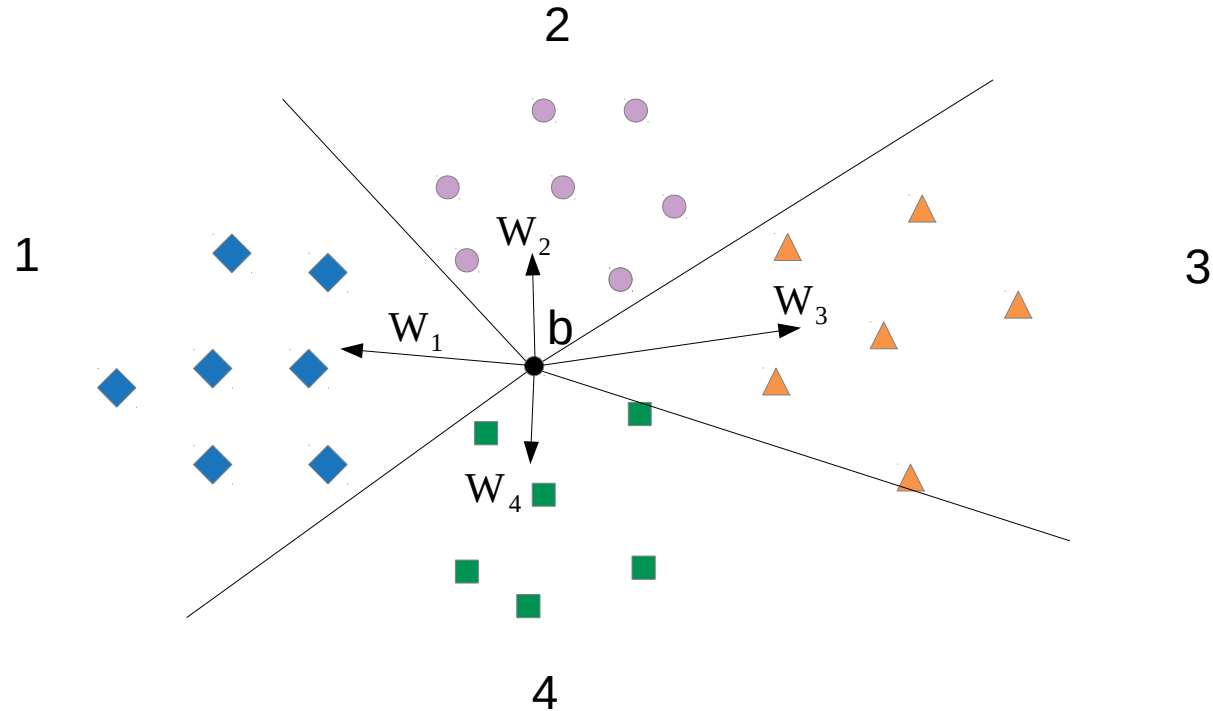- Parameters: $W \in \mathbb{R}^{k \times d}, b \in \mathbb{R}^k$

Logits

$$s = W x + b$$

$$\hat{y} = \text{argmax}_i s_i \qquad P(y) = \text{softmax}(s)_y$$
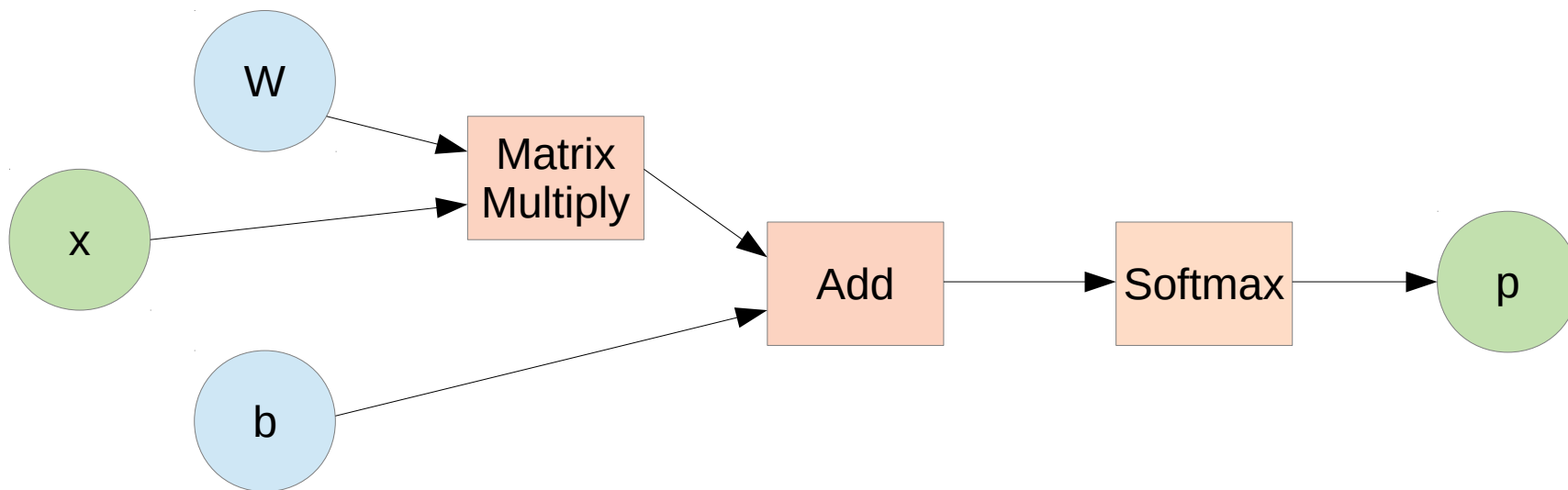
Linear Classification

Softmax

- Turn scores into a distribution

$$P(y) = \text{softmax}(\text{s})_y = \frac{e^{s_y}}{\sum_i e^{s_i}}$$

Loss: $-\log P(y)$

Computation Graphs

- Arrange the operations of a network as a graph

Computation Graphs - Abstraction

- Group tensors and operations into repeatable layers